identifying documents relating to the query by comparing search terms in the query to an index of a corpus;

generating a plurality of multiword substrings from the query in which each of the substrings includes at least two words;

calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substring; and

selecting semantic units from the generated multiword substrings based on the calculated values.

5. (Amended) The method of claim 1, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the identified documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents.

6. (Amended) A method of locating documents in response to a search query, the method comprising:

receiving the search query from a user;

generating a list of relevant documents based on search terms of the query;

identifying a subset of documents that are most relevant ones of the documents in the list of relevant documents;

generating a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words;

calculating, for each of the generated substrings, a value related to one or more documents in the subset of documents that contain the substring;

selecting semantic units from the generated multiword substrings based on the calculated values; and

refining the generated list of relevant documents based on the selected semantic units.

10. (Amended) The method of claim 6, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents.

11. (Amended) A system comprising:

a server connected to a network, the server receiving search queries from users via the network, the server including:

at least one processor; and

a memory operatively coupled to the processor, the memory storing program instructions that when executed by the processor, cause the processor to: identify a list of documents relating to the search query by matching individual search terms in the query to an index of a corpus; generate a plurality of multiword substrings from the query in which each of the substrings includes at least two words; calculate, for each of the generated substrings, a value relating to one or more documents of the

identified list of documents that contain the generated substring; and select semantic units from the generated multiword substrings based on the calculated values.

17. (Amended) The system of claim 11, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur in less relevant documents.

18. (Amended) A server comprising:

a processor; and

a memory operatively coupled to the processor, the memory including:

a ranking component configured to return a list of documents ordered by relevance in response to a search query; and

a semantic unit locator component configured to locate semantic units, each having a plurality of words, in search queries entered by a user based on a predetermined number of most relevant documents in the list of documents returned by the ranking component.

24. (Amended) The server of claim 21, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur in less relevant documents.

25. (Amended) A computer-readable medium storing instructions for causing at least one processor to perform a method that identifies semantic units within a search query, the method comprising:

identifying documents relating to the query by matching individual search terms in the query to an index of a corpus;

A5

forming a plurality of multiword substrings of the query in which each of the substrings includes at least two words;

calculating, for each of the substrings, a value relating to the portion of the identified documents that contain the substring; and

selecting semantic units from the generated multiword substrings based on the calculated values.

29. (Amended) The computer-readable medium of claim 27, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned

A6

higher calculated values than substrings that occur in less relevant documents.

30. (Amended) A computer-readable medium storing instructions for causing a processor to perform a method, the method comprising:

receiving the search query from a user;

generating a list of relevant documents based on individual search terms of the query;

identifying a subset of documents that are the most relevant documents from the list of relevant documents;

forming a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words;

A6

calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring;

selecting semantic units from the generated multiword substrings based on the calculated values; and

refining the generated list of relevant documents based on the selected semantic units.

34. (Amended) The computer-readable medium of claim 30, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur in less relevant documents.

A7

36. (Amended) An apparatus for locating documents in response to a search query, comprising:

A8

means for receiving the search query from a user;

means for generating a list of relevant documents based on individual search terms of the query;

means for identifying a subset of documents that are the most relevant documents from the list of relevant documents;

means for forming a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words;

means for calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring;

means for selecting semantic units from the generated multiword substrings based on the calculated values; and

means for refining the generated list of relevant documents based on the selected semantic units.

37. (New Claim) The method of claim 1, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents

38. (New Claim) The method of claim 6, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that an occurrence of a substring in a more relevant one of the identified documents is weighted more than an occurrence of the substring in a less relevant one of the documents

39. (New Claim) The system of claim 11, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that